

Document de recherche

Direction des études analytiques : documents de recherche

Augmentation des gains des immigrants : biais de sélection ou progrès réel?

par Garnett Picot et Patrizio Piraino

Division de l'analyse sociale
24-I, Immeuble R.-H.-Coats, 100, promenade Tunney's Pasture
Ottawa, Ontario K1A 0T6

Téléphone: 1-800-263-1136



Statistique
Canada

Statistics
Canada

Canada

Augmentation des gains des immigrants : biais de sélection ou progrès réel?

par
Garnett Picot et Patrizio Piraino

11F0019M N° 340
ISSN 1205-9161
ISBN 978-1-100-98811-5

Statistique Canada
Division de l'analyse sociale
24-I, Immeuble R.-H.-Coats, 100 promenade Tunney's Pasture, Ottawa K1A 0T6

Comment obtenir d'autres renseignements :
Service national de renseignements : 1-800-263-1136
Renseignements par courriel : infostats@statcan.gc.ca

Février 2012

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2012

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada.
(<http://www.statcan.gc.ca/reference/copyright-droit-auteur-fra.htm>)

The English version of this publication is available (catalogue no. 11F0019M, no. 340).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées dans le site www.statcan.gc.ca sous « Notre organisme » cliquez sur À propos de nous > Notre organisme > et sélectionnez « Offrir des services aux Canadiens ».

Études analytiques Documents de recherche

La série de documents de recherche de la Direction des études analytiques permet de faire connaître, avant leur publication, les travaux de recherche effectués par le personnel de la Direction, les boursiers invités et les universitaires associés. Cette série a pour but de favoriser la discussion sur divers sujets, notamment le travail, la dynamique des entreprises, les pensions, l'agriculture, la mortalité, la langue, l'immigration, l'informatique statistique et la simulation. Le lecteur est invité à faire part aux auteurs de ses commentaires, critiques et suggestions. La liste des titres figure à la fin du document.

Les documents de la série sont distribués aux établissements de recherche et aux bibliothèques spécialisées. Ces documents peuvent être téléchargés à partir d'Internet, à l'adresse www.statcan.gc.ca.

Comité de révision des publications
Études analytiques, Statistique Canada
Immeuble R.-H.-Coats, 24^e étage
Ottawa (Ontario) K1A 0T6

Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- ^p provisoire
- ^r révisé
- x confidentiel en vertu des dispositions de la Loi sur la statistique
- E à utiliser avec prudence
- F trop peu fiable pour être publié
- * valeur significativement différente de l'estimation pour la catégorie de référence ($p < 0,05$)

Table des matières

Résumé.....	5
Sommaire	6
1 Introduction	8
2 L'enjeu : l'utilisation de données transversales répétées pour estimer le biais dans l'augmentation des gains des immigrants	10
2.1 Résultats empiriques d'études antérieures.....	11
3 Données.....	13
4 Résultats empiriques	15
4.1 Estimations par régression.....	18
5 Qui sort de l'échantillon?	23
6 Conclusions.....	27
7 Annexe	28
Bibliographie.....	32

Résumé

Le présent document porte sur l'effet de l'attrition sélective sur les estimations de l'augmentation des gains des immigrants, à partir de données transversales répétées au Canada. Des observations récentes à partir de données longitudinales aux États-Unis montrent que l'écart entre les gains des immigrants et ceux des personnes nées aux États-Unis diminue plus lentement au fil des années suivant l'établissement, que ne le laissent supposer les estimations transversales antérieures. Cela vient du fait que les résultats fondés sur des données transversales répétées comprenaient un biais découlant de l'attrition sélective des immigrants. La présente étude utilise des données fiscales longitudinales couplées aux fiches relatives au droit d'établissement des immigrants afin d'estimer la variation des gains des immigrants et l'écart entre les gains des immigrants et ceux des personnes nées au Canada. Les résultats sont comparés à ceux découlant de données transversales répétées. Lorsque l'on se penche sur l'augmentation des gains des immigrants, on remarque que les trajectoires des gains fondées sur des sections transversales répétées sont légèrement biaisées à la hausse par suite de l'attrition sélective des immigrants. Toutefois, il n'y a pas de preuve de biais dans la trajectoire de l'écart entre les gains des immigrants et ceux des personnes nées au Canada sur la base des résultats obtenus au moyen de données transversales répétées au Canada. Même si les immigrants à faibles gains sont plus susceptibles que leurs homologues à gains élevés de sortir des échantillons transversaux au fil du temps, il en va de même pour la population née au Canada. Ainsi, aucune preuve de biais n'est observée lorsque l'on compare les trajectoires des gains des immigrants avec celles des personnes nées au Canada.

Classifications du JEL : J31, J61

Mots-clés : immigration, assimilation, données longitudinales, biais de sélection

Sommaire

L'écart entre les gains des immigrants et ceux de travailleurs comparables nés au pays est peut-être le sujet le plus étudié dans le domaine de l'économie de l'immigration. Des recherches de cette nature devraient idéalement être fondées sur des données longitudinales permettant de suivre les gains des immigrants après leur arrivée, au fur et à mesure qu'ils s'établissent au Canada. Toutefois, la plupart des recherches existantes au Canada sur les trajectoires des gains des immigrants sont fondées sur des données transversales du Recensement de la population, et non pas sur des données longitudinales. Il y a deux explications à cela : les données longitudinales ne sont disponibles que depuis peu de temps et elles comprennent relativement peu de covariables socioéconomiques. De façon plus particulière, dans la plupart des données longitudinales canadiennes, il n'est pas possible de tenir compte des différences de scolarité entre les immigrants et les personnes nées au Canada.

Généralement, les chercheurs se tournent vers des données transversales répétées du recensement pour élaborer des cohortes pseudo-longitudinales de personnes. Les immigrants qui sont arrivés au Canada au cours de la période de 1991 à 1995 sont dénombrés par le Recensement de 1996, de 1 à 5 ans après leur établissement. Les immigrants de la même cohorte qui restent au Canada sont dénombrés par le Recensement de 2001, de 6 à 10 ans après leur établissement, etc. De cette façon, on a pu estimer à la fois l'augmentation des gains des immigrants et la variation de l'écart entre la rémunération des immigrants et celle des personnes nées au Canada pour diverses cohortes.

Toutefois, la composition des cohortes pseudo-longitudinales change au fil du temps, au fur et à mesure que les immigrants quittent leur pays hôte ou leur emploi. Cela peut causer un biais dans les trajectoires des gains estimées à partir de données transversales. Si, par exemple, les immigrants qui sortent de l'échantillon sont plus susceptibles d'avoir de moins bons résultats sur le marché du travail que ceux qui y restent (et sont donc incités à partir), la trajectoire des gains fondée sur des cohortes pseudo-longitudinales sera biaisée à la hausse. Au fur et à mesure que le temps passe, toute cohorte pseudo-longitudinale comprendra de plus en plus d'immigrants de la cohorte originale « qui réussissent », c'est-à-dire ceux qui ont des gains plus élevés. Ainsi, une part importante de l'augmentation des gains au fil du temps (c.-à-d. en fonction du nombre d'années écoulées depuis l'établissement) peut être le résultat d'un changement dans la composition de la cohorte (une forme de biais de sélection de l'échantillon), et non celui d'une augmentation réelle des gains. Les recherches menées aux États-Unis sur les gains des immigrants ont permis de déceler un tel biais.

Le présent document repose sur des données longitudinales tirées de dossiers administratifs, couplées aux fiches relatives au droit d'établissement des immigrants. L'augmentation des gains des immigrants et la différence entre les gains des immigrants et ceux des personnes nées au Canada sont estimées pour les années écoulées depuis la migration, pour trois cohortes d'immigrants qui sont arrivés depuis le début des années 1980. Ces données permettent d'estimer les résultats au chapitre des gains, grâce à l'utilisation de véritables données longitudinales et de cohortes pseudo-longitudinales établies à partir de sections transversales répétées de la même source de données. Cette approche élimine les différences entre les résultats, qui pourraient découler des différents modes de collecte et procédures s'appliquant aux divers ensembles de données. Afin d'établir un lien plus étroit entre les résultats de l'étude et ceux des ouvrages publiés antérieurement, les trajectoires des gains des immigrants en fonction du nombre d'années écoulées depuis la migration sont aussi estimées à partir de sections transversales répétées du recensement canadien. Les résultats des deux sources de données transversales (le recensement et les données administratives) sont alors comparés à ceux tirés d'une source de données longitudinales véritables (les données administratives), afin de déterminer s'il existe une preuve de

biais dans la baisse de l'écart entre les gains des immigrants et ceux des personnes nées au Canada ou dans les trajectoires des gains des immigrants seulement.

L'analyse fournit peu de preuves d'un biais dans l'écart entre les gains des immigrants et ceux des personnes nées au Canada calculé à partir de sections transversales répétées, comparativement à des données longitudinales véritables. Même si les immigrants moins bien rémunérés des diverses cohortes sont plus susceptibles de sortir de l'échantillon que leurs homologues mieux rémunérés, il semble en être de même pour les personnes nées au Canada. C'est donc dire que l'augmentation des gains des cohortes d'immigrants et des cohortes de personnes nées au Canada est surestimée à peu près dans la même mesure dans les données transversales, ce qui explique que la trajectoire de l'« écart » obtenue par l'estimation du modèle standard d'assimilation à partir de données longitudinales ne fait ressortir qu'un faible biais dans les études antérieures de l'assimilation des gains au Canada. Ce résultat contraste avec certains résultats obtenus aux États-Unis. Toutefois, lorsque l'on se penche sur l'augmentation des gains des immigrants et de celle des gains des personnes nées au Canada séparément, plutôt que sur l'écart entre les gains des deux groupes, la recherche laisse supposer que les trajectoires des gains fondées sur des sections transversales répétées sont biaisées à la hausse pour les deux groupes, par suite de l'attrition sélective.

1 Introduction

Au cours de leurs toutes premières années dans un pays hôte, les immigrants gagnent généralement moins que les personnes nées au pays. Toutefois, les immigrants voient leurs gains relatifs augmenter les années subséquentes, au fur et à mesure qu'ils acquièrent de l'expérience dans le pays hôte, des connaissances linguistiques et des connaissances des marchés du travail locaux (Chiswick, 1978; Meng, 1987; Borjas, 1999). Par exemple, au cours de leurs cinq premières années au Canada, les immigrants qui se sont établis à la fin des années 1970 avaient des gains représentant 85 % de ceux de leurs homologues nés au Canada; après 11 à 15 ans, leurs gains relatifs avaient atteint 92 %. Les gains des cohortes d'immigrants plus récentes étaient généralement plus faibles. Les gains des immigrants qui se sont établis au début des années 1990 représentaient 60 % de ceux de leurs homologues nés au Canada au cours de leurs cinq premières années au Canada, atteignant 78 % après 11 à 15 ans (Frenette et Morissette, 2003)¹.

Les trajectoires des gains des immigrants, ainsi que les différences entre les trajectoires des gains des immigrants des cohortes d'établissement successives, sont plus faciles à étudier au moyen de données longitudinales comprenant des tailles d'échantillons importantes (afin de distinguer les différentes cohortes d'établissement), ainsi que de données sur les caractéristiques socioéconomiques (pour tenir compte des différences entre les populations immigrantes et celles nées au pays). Même si certaines études récentes reposent sur des données administratives longitudinales (Hu, 2000; Edin et coll., 2000; Duleep et Dowhan, 2002; Green et Worswick, 2004; Lubotsky, 2007; Aydemir et Robinson, 2008), la plupart des recherches sur les gains initiaux des immigrants et les trajectoires des gains subséquents sont fondées sur les données du recensement. Cela vient du fait que des données longitudinales ne sont disponibles que depuis peu et comprennent relativement peu de covariables socioéconomiques. Plus particulièrement, les données longitudinales canadiennes ne permettent pas aux chercheurs de tenir compte des différences de scolarité entre les immigrants et les personnes nées au Canada.

Généralement, les chercheurs ont recours à des données transversales répétées du Recensement de la population au Canada pour établir des cohortes pseudo-longitudinales d'immigrants. Par exemple, les immigrants établis au Canada entre 1991 et 1995 sont dénombrés par le Recensement de 1996, de 1 à 5 ans après leur arrivée. Les immigrants de cette cohorte qui restent au Canada seront dénombrés par le Recensement de 2001, de 6 à 10 ans après leur établissement; par le Recensement de 2006, de 11 à 15 ans après leur établissement, etc. Sur cette base, l'augmentation des gains et les variations de l'écart entre la rémunération des immigrants et celle des personnes nées au Canada ont été estimées pour diverses cohortes d'immigrants.

Toutefois, les échantillons de ces panels de cohortes pseudo-longitudinales changent au fil du temps, certains immigrants quittant le pays hôte. Pour le Canada, Aydemir et Robinson (2008) se sont penchés sur les jeunes immigrants de sexe masculin, un groupe très mobile, et ont estimé qu'environ le tiers d'entre eux sont partis au cours de leurs 20 premières années au Canada, plus de la moitié du groupe partant pendant la première année. Bien que, globalement, les taux de départ chez les cohortes d'immigrants sont probablement plus faibles, le départ des immigrants peut causer un biais dans les trajectoires des gains estimées à partir de données transversales. Si, par exemple, les immigrants qui partent ont tendance à avoir de moins bons résultats sur le marché du travail que ceux qui restent (et sont par conséquent incités à partir), la trajectoire des gains fondée sur des cohortes pseudo-longitudinales établies à partir de données transversales sera biaisée à la hausse. Au fur et à mesure que le temps passe, la cohorte pseudo-longitudinale

1. Plus précisément, ces chiffres représentent le logarithme de rapports des gains (gains des immigrants et gains des personnes nées au Canada). Un certain nombre d'études ont porté sur la diminution des gains relatifs initiaux des cohortes d'entrée successives d'immigrants au Canada (Bloom et Gunderson, 1991; McDonald et Worswick, 1998; Baker et Benjamin, 1994; et Grant, 1999). Picot et Sweetman (2005) se penchent sur cette question.

comprendra de plus en plus d'immigrants « qui réussissent », c'est-à-dire ceux qui ont des gains plus élevés. Ainsi, une part importante des augmentations des gains des immigrants au fur et à mesure des années écoulées depuis l'année de l'immigration peut être le résultat d'un changement dans la composition de la cohorte, une forme de biais de sélection de l'échantillon, et non pas le résultat d'une augmentation réelle des gains.

C'est exactement le résultat auquel arrivent Hu (2000) et Lubotsky (2007) aux États-Unis. De façon plus particulière, l'écart entre les gains des immigrants et ceux des personnes nées au pays diminue deux fois plus rapidement lorsqu'il est estimé au moyen de cohortes pseudo-longitudinales établies à partir de données transversales répétées du recensement aux États-Unis, que lorsqu'il est estimé au moyen de données longitudinales véritables. Lubotsky conclut que la probabilité plus forte de migration de sortie chez les immigrants à faible rémunération a mené systématiquement les chercheurs par le passé à surestimer l'augmentation de la rémunération des immigrants qui restent aux États-Unis. Ces constatations donnent un aperçu moins optimiste des résultats économiques obtenus au fil du temps par les immigrants aux États-Unis.

Les intervenants dans le débat stratégique sur l'immigration aux États-Unis citent souvent le système canadien comme un système dans lequel l'immigration des travailleurs hautement qualifiés est activement encouragée. Il reste à savoir si un système de points similaire à celui utilisé au Canada améliorerait les résultats économiques des immigrants aux États-Unis. Toutefois, le fait de déterminer si l'augmentation des gains des immigrants au Canada est surestimée, comme cela semble être le cas aux États-Unis, contribuera à éclairer les décideurs des deux pays.

Les données utilisées dans la présente étude sont décrites de façon détaillée à la section 3. Essentiellement, des données longitudinales fondées sur les déclarations de revenus annuelles des particuliers sont couplées aux fiches relatives au droit d'établissement des immigrants. Cela produit un large échantillon représentatif de travailleurs, qui inclut à la fois des immigrants et des personnes nées au Canada. L'augmentation des gains des immigrants et la différence entre les gains des immigrants et ceux des personnes nées au Canada sont estimées en fonction des années écoulées depuis la migration, pour trois cohortes d'immigrants qui sont arrivées depuis le début des années 1980. Ces données permettent d'estimer les résultats au chapitre des gains, grâce à l'utilisation de données longitudinales véritables et de cohortes pseudo-longitudinales établies à partir de sections transversales répétées de la même source de données. Il est important que l'on puisse obtenir des données transversales et longitudinales à partir de la même source de données pour éliminer les différences entre les résultats possiblement causés par des modes de collecte et des procédures qui sont dissemblables d'un ensemble de données à l'autre. Afin de pouvoir lier plus étroitement ces résultats aux ouvrages publiés, on estime aussi les trajectoires des gains des immigrants au moyen de cohortes pseudo-longitudinales établies à partir de sections transversales répétées du recensement canadien.

L'analyse fournit peu de preuves d'un biais dans l'écart entre les gains des immigrants et ceux des personnes nées au Canada calculé à partir de données transversales répétées. Même si les immigrants moins bien rémunérés des trois cohortes sont plus susceptibles de sortir de l'échantillon transversal, il en va de même pour les personnes nées au Canada. C'est donc dire que l'augmentation des gains des cohortes d'immigrants et de personnes nées au Canada est surestimée dans les données transversales, à peu près dans la même mesure. Ainsi, la trajectoire de l'écart entre les gains estimée au moyen d'un modèle standard de gains et de données longitudinales indique qu'il existe peu de biais dans les résultats des études antérieures sur les gains relatifs des immigrants au Canada. Ce résultat contraste avec ceux obtenus aux États-Unis.

Le reste du document se répartit ainsi. La section 2 explique le biais possible dans l'augmentation des gains des immigrants et passe en revue le petit nombre d'ouvrages empiriques sur le sujet. La section 3 décrit les données utilisées dans l'étude, y compris leurs forces et leurs faiblesses. Les

résultats empiriques sont présentés et examinés dans les sections 4 et 5. La conclusion figure à la section 6.

2 L'enjeu : l'utilisation de données transversales répétées pour estimer le biais dans l'augmentation des gains des immigrants

En ce qui a trait à l'estimation de l'augmentation des gains des immigrants, la situation idéale en serait une où tous les immigrants des cohortes d'établissement successives resteraient au Canada au cours d'une période de temps prolongée (p. ex., 20 ans) et où des données longitudinales sur les gains et les caractéristiques socioéconomiques seraient disponibles pour chaque cohorte. Dans ce cas, la production d'estimations non biaisées de l'augmentation des gains serait simple. Toutefois, la situation est plus complexe. Certains immigrants partent du Canada et, après une période prolongée, un sous-échantillon d'immigrants de « longue date » restent. La composition de ce sous-échantillon peut différer de celle de la cohorte d'établissement initiale, particulièrement si les immigrants qui sont partis sont ceux qui ont obtenu les moins bons résultats. Il existe une autre considération, à savoir que les chercheurs n'ont peut-être pas de données longitudinales et utilisent peut-être à la place des données transversales répétées pour établir des cohortes pseudo-longitudinales d'immigrants. La question du biais de composition se pose à nouveau.

Le présent document, qui porte particulièrement sur un échantillon d'immigrants de longue date, estime l'augmentation des gains au moyen de données longitudinales dont la composition d'échantillon est fixe, ainsi que de données transversales répétées dont la composition d'échantillon est variable. En comparant les résultats des deux approches, on remarque que l'on peut déterminer dans quelle mesure les estimations fondées sur des données transversales répétées sont touchées par un biais de composition.

Le présent document ne porte pas sur le biais découlant du fait que l'échantillon ne comprend pas tous les immigrants qui sont arrivés au Canada dans une cohorte particulière, mais se limite plutôt aux immigrants de longue date. Toutefois, c'est le biais possible découlant des changements de composition des cohortes qui préoccupe principalement les chercheurs.

La figure 1 explique la différence dans la mesure de l'augmentation des gains des immigrants à partir de données longitudinales et de données transversales répétées, dans le contexte du recensement canadien. Les lignes de la figure indiquent l'année d'arrivée de trois cohortes d'immigrants (1985, 1990, 1995), tandis que les colonnes montrent l'année pendant laquelle les gains ont été mesurés. Dans chaque cellule, $E(w)$ représente les gains moyens mesurés au temps c (colonne) pour la cohorte des immigrants qui sont arrivés au Canada au temps r (ligne).

Dans la partie A de la figure 1, on voit que les échantillons transversaux produiront des estimations de l'augmentation des gains des immigrants qui dépendront du taux de sortie des immigrants de l'échantillon. Par exemple, la première ligne montre que les gains pour la cohorte de 1985 seront mesurés en 1990, les gains moyens étant calculés pour le sous-ensemble des immigrants toujours compris dans l'échantillon après cinq ans. En 2005, les gains moyens de la même cohorte seront calculés pour le sous-ensemble des immigrants qui sont toujours dans l'échantillon après 20 ans. Lorsque l'on examine les différentes années, on note que chaque immigrant contribuera au taux estimé d'augmentation des gains de la cohorte d'établissement, tant et aussi longtemps qu'il restera dans l'échantillon. Si ceux qui partent ont tendance à constituer un sous-échantillon non aléatoire de ceux qui sont arrivés initialement, il se produira un biais de composition dans la trajectoire des gains estimée.

Dans la partie B de la figure 1 (données longitudinales), l'échantillon se limite aux immigrants qui sont compris dans la première année de données. Pour chaque cohorte d'immigrants, cela permet d'estimer les gains moyens pour les immigrants qui sont restés dans l'échantillon², ce qui produit une estimation non biaisée de l'augmentation des gains pour ce sous-ensemble de personnes³.

Figure 1

Mesures des gains moyens des immigrants, données longitudinales et données transversales

Année d'arrivée	Année d'observation			
	1990	1995	2000	2005
Partie A – Section transversale répétée				
1985	E(w 5 ans)	E(w 10 ans)	E(w 15 ans)	E(w 20 ans)
1990	...	E(w 5 ans)	E(w 10 ans)	E(w 15 ans)
1995	E(w 5 ans)	E(w 10 ans)
Partie B – Données longitudinales				
1985	E(w 20 ans)	E(w 20 ans)	E(w 20 ans)	E(w 20 ans)
1990	...	E(w 15 ans)	E(w 15 ans)	E(w 15 ans)
1995	E(w 10 ans)	E(w 10 ans)

Note : E(w) représente les gains moyens mesurés au temps c (années figurant dans les colonnes) pour la cohorte des immigrants qui sont arrivés au Canada au temps r (années figurant aux lignes).

À noter que les immigrants à faibles gains n'ont pas à émigrer de façon disproportionnée d'un pays pour que ce biais se produise et que les données sur les taux d'émigration ne sont pas nécessaires pour estimer le biais. La principale préoccupation a plutôt trait au nombre disproportionné d'immigrants à faibles gains qui quittent leur emploi (plutôt que ceux qui partent du fait de l'émigration), et l'échantillon d'intérêt est constitué de la population active occupée. En fait, lorsque l'on estime des trajectoires des gains des immigrants au moyen de données transversales répétées du recensement, seules les observations comportant des gains positifs dans chaque section transversale sont utilisées généralement.

2.1 Résultats empiriques d'études antérieures

Il existe très peu d'ouvrages publiés permettant de déterminer si la migration de sortie sélective des immigrants donne lieu à un biais dans les estimations transversales des résultats économiques des immigrants. Sur la base des données américaines, Hu (2000) et, plus récemment, Lubotsky (2007) ont conclu que l'émigration sélective donne lieu à une surestimation des résultats économiques des immigrants⁴. De façon plus particulière, Lubotsky a utilisé les données sur les gains longitudinaux des dossiers de la sécurité sociale pour la période de 1951 à 1997 et a démontré que l'écart entre les gains des immigrants et ceux des personnes nées au pays diminue deux fois moins rapidement dans les données longitudinales que dans les données

2. Cela n'exclut pas la possibilité qu'ils soient absents certaines années.

3. Cela n'est pas équivalent à l'estimation de l'augmentation des gains de la cohorte d'arrivée, si toutes les personnes étaient restées jusqu'en 2005. Cette estimation pourrait être obtenue à partir des données longitudinales sur les personnes qui sont restées, uniquement si on souhaitait partir du principe que l'attrition est fondée sur des attributs permanents qui ne sont pas liés à l'augmentation des gains des immigrants au fil du temps. Cette interprétation est laissée de côté, le présent document visant principalement à vérifier si les estimations existantes de l'augmentation des gains des immigrants obtenues au moyen de sections transversales répétées du recensement sont biaisées par suite du changement dans la composition de la cohorte d'immigrants dans les données transversales. C'est cette question qui a préoccupé les chercheurs qui utilisent les données transversales et celle sur laquelle la recherche américaine a été axée (p. ex., Lubotsky 2007).

4. Edin et coll. (2000) arrivent à une conclusion similaire dans leur analyse des données suédoises.

transversales répétées découlant du recensement décennal aux États-Unis. Comme l'indique Lubotsky, toutefois, le « biais » n'est pas uniforme pour toutes les cohortes d'entrée, ressortant davantage parmi la cohorte arrivée au cours des années 1970 et seulement de façon marginale parmi les cohortes arrivées dans les années 1960 et 1980.

Deux documents utilisent les données canadiennes pour étudier cette question, mais de façon moins directe que dans les recherches américaines, où les résultats des données longitudinales et transversales sont comparés directement. Les deux documents canadiens sont fondés sur l'Enquête sur la dynamique du travail et du revenu (EDTR), un panel longitudinal sur 6 ans de travailleurs canadiens dans lequel on peut identifier les immigrants. Hum et Simpson (2000) ont examiné l'augmentation des gains sur la période de 1993 à 1997 et ont trouvé que, même dans les données longitudinales non corrigées, l'écart entre la rémunération des immigrants et celle des hommes nés au Canada variait peu. L'augmentation des gains était à peu près la même pour les deux groupes au cours de la période d'étude de cinq ans. Chez les femmes, on a noté une augmentation, plutôt qu'une diminution, de l'écart de rémunération non corrigé, l'augmentation des gains ayant été plus forte chez les femmes nées au Canada que les immigrantes. Par ailleurs, à partir d'un modèle à effets fixes, Hum et Simpson concluent qu'il n'y a pas de preuve de rétrécissement de l'écart entre la rémunération des immigrants et celle des hommes nés au Canada⁵. Cela contraste avec à peu près toutes les études canadiennes existantes fondées sur des données transversales répétées du recensement. Hum et Simpson concluent que leurs résultats montrent que les constatations découlant des données transversales, qui peuvent être sujettes à un biais découlant de l'hétérogénéité nonobservée des travailleurs, devraient être interprétées avec précaution.

Dans une étude plus récente, Skuterud et Su (2009) ont regroupé quatre panels de données de l'EDTR recueillies entre 1993 et 2004, afin d'augmenter l'échantillon longitudinal d'immigrants et de personnes nées au Canada. Contrairement à Hum et Simpson (2000), ils ont déterminé que les immigrants avaient connu des gains économiques considérables. De façon plus pertinente dans le contexte du présent examen, Skuterud et Su ont aussi tenté de résoudre le problème de biais dans l'augmentation de la rémunération des immigrants. Comme les panels de leurs données sont assez limités, ils ont utilisé une approche substantiellement différente de celle utilisée dans le présent document ou par Lubotsky (2007). De façon plus particulière, Skuterud et Su ont utilisé un modèle à effets fixes pour éliminer, dans la mesure du possible, l'effet des caractéristiques individuelles non observées sur l'augmentation de l'émigration et de la rémunération (c.-à-d. les effets de la migration de sortie sélective sur l'augmentation de la rémunération). Ils ont conclu que l'approche à effets fixes modifiait relativement peu les estimations de l'augmentation de la rémunération et ne signifiait pas une augmentation substantiellement plus faible de la rémunération des immigrants dans les données longitudinales comme les ouvrages publiés aux États-Unis ont eu tendance à le démontrer (p. ex., Lubotsky, 2007). Leurs résultats font ressortir qu'il est possible que la nature de la migration de sortie diffère au Canada et aux États-Unis et qu'on ne devrait pas s'attendre à un biais à la hausse dans les estimations transversales existantes des résultats économiques des immigrants.

La présente étude, qui profite de données administratives de plus grande qualité, peut aider à jeter de la lumière sur ces résultats différents pour le Canada. Par ailleurs, compte tenu des périodes de référence plus longues utilisées, l'effet des « sorties » sélectives peut être examiné et comparé avec les résultats des recherches effectuées aux États-Unis. Pour ce faire, on adopte la même approche analytique que celle suivie par Hu (2000) et Lubotsky (2007), qui consiste à sélectionner des échantillons d'immigrants en fonction du nombre d'années écoulées depuis la migration et à examiner les trajectoires des gains sur cette période.

5. Cette constatation est confirmée dans une étude ultérieure qui est fondée sur le même ensemble de données (Hum et Simpson, 2004).

3 Données

La présente étude utilise trois sources de données : la banque de Données administratives longitudinales (DAL); la Base de données longitudinales sur les immigrants (BDIM) et le Recensement de la population. La banque DAL représente un sous-échantillon aléatoire de 20 % du fichier sur la famille T1 (FFT1). T1 correspond à la déclaration de revenu générale T1 (*Déclaration de revenus et de prestations*) des particuliers. Le FFT1 est un fichier transversal annuel de tous les déclarants fiscaux au Canada et de leurs familles. Même si une personne doit produire une déclaration de revenus individuelle pour être comprise dans le FFT1 (et par conséquent dans la banque DAL), la couverture de la population est très élevée, soit environ 95 % de la population en âge de travailler. Cela vient du fait que les réductions d'impôt encouragent les personnes qui ont un faible revenu imposable ou qui n'ont aucun revenu imposable à produire une déclaration. Les dossiers d'une personne figurant dans la banque DAL sont couplés d'une année à l'autre au moyen d'un numéro d'identification unique, ce qui permet la création d'un profil individuel. Outre les données sur les gains annuels, la banque DAL comprend des renseignements sur la date de naissance et le sexe de chaque personne⁶.

La BDIM est constituée des fiches d'établissement des immigrants et des dossiers de l'impôt sur le revenu des particuliers T1. Les premières fournissent des renseignements sur les caractéristiques des immigrants, tandis que les seconds comportent des renseignements longitudinaux sur les caractéristiques financières, les gains d'emploi ayant la plus grande pertinence pour la présente étude. Compte tenu de la couverture presque universelle des dossiers fiscaux, la BDIM permet de suivre les trajectoires des gains des cohortes d'immigrants du début des années 1980 jusqu'en 2005. Dans le présent document, on utilise un ensemble de données couplées de la banque DAL et de la BDIM. Le couplage est possible parce que l'identificateur individuel unique est disponible dans les deux ensembles de données.

L'analyse empirique porte sur trois cohortes d'immigrants successives : ceux qui se sont établis au Canada de 1985 à 1989, de 1990 à 1994 et de 1995 à 1999. Étant donné que les données sont disponibles jusqu'en 2005, les trajectoires des gains des trois cohortes peuvent être suivies pendant 20, 15 et 10 ans après l'établissement, respectivement. Les gains des immigrants au fil du temps sont analysés à la fois en termes absolus et par rapport à ceux des personnes nées au Canada (c.-à-d. l'écart entre les gains des immigrants et ceux des personnes nées au Canada).

Comme dans un certain nombre d'études antérieures, l'analyse est axée sur les hommes, afin d'éviter les complications découlant d'une activité sélective sur le marché du travail. Les immigrants sont définis comme des personnes nées à l'étranger qui avaient entre 25 et 44 ans au moment de leur établissement au Canada, comme en fait foi la fiche relative au droit d'établissement⁷. Pour produire des trajectoires des gains des hommes nés au Canada qui correspondent à celles des immigrants dans l'étude, des groupes de comparaison des personnes nées au Canada ont été constitués à partir des mêmes cohortes de naissance que celles utilisées pour les immigrants. Enfin, l'ensemble de l'échantillon se limite aux observations d'années-personnes pour les 25 à 64 ans.

6. La définition des gains comprend les traitements, salaires et commissions, avant retenues, ainsi que les revenus d'emploi imposables autres que les traitements, salaires et commissions (p. ex., les pourboires, gratifications ou allocations de présence). Elle exclut le revenu d'un travail autonome. De plus amples détails sur l'ensemble de données sont disponibles dans Statistique Canada (2009).

7. Les immigrants qui se situaient à l'extérieur de cette fourchette d'âge au moment de l'établissement, de même que les travailleurs étrangers temporaires, sont supprimés de l'analyse. La limite d'âge inférieure est imposée parce que l'expérience sur le marché du travail des très jeunes immigrants s'apparentera probablement davantage à celle des travailleurs nés au Canada qu'à celles des immigrants adultes. La limite d'âge supérieure sert à diriger l'attention sur les immigrants dont l'immigration remonte à plus loin. La sensibilité des résultats principaux à cette restriction est vérifiée en annexe.

Un aspect utile de la DAL-BDIM est qu'elle permet le tirage d'échantillons « transversaux » et « longitudinaux ». Comme la source de données est mise à jour chaque année au moyen de nouvelles observations, les fichiers annuels demeurent représentatifs au niveau transversal (Statistique Canada, 2009). Des échantillons transversaux de toutes les observations d'années-personnes comportant des gains positifs ont été sélectionnés pour 1990, 1995, 2000 et 2005 — toutes des années de recensement⁸. Les échantillons transversaux regroupés ont été par la suite utilisés pour élaborer des cohortes pseudo-longitudinales, selon une méthode correspondant aux études antérieures fondées sur le recensement et traitant des gains des immigrants.

L'échantillon longitudinal est constitué des données sur les gains annuels des hommes de chaque cohorte d'immigrants et de chaque cohorte de personnes nées au Canada pour toutes les années disponibles. Le critère de sélection essentiel pour l'inclusion dans l'échantillon longitudinal est le fait pour la personne d'avoir des gains positifs pour l'année de données la plus récente. Il s'agit de la définition appropriée, l'objectif étant d'évaluer le biais dans les estimations transversales de l'augmentation des gains des immigrants. En fait, les estimations de l'assimilation des gains des immigrants découlant de vagues regroupées du recensement sont fondées sur une restriction de gains positifs pour chaque section transversale utilisée.

Un échantillon pseudo-longitudinal est aussi élaboré à partir des fichiers transversaux des recensements de 1990, 1995, 2000 et 2005⁹. Cela permet la comparaison des estimations de l'augmentation des gains fondées sur les données administratives et les données de recensement. Pour plus d'uniformité avec l'échantillon administratif, seuls les hommes âgés de 25 à 64 ans ayant des gains positifs sont inclus¹⁰.

Plusieurs caractéristiques des données comportent des avantages par rapport aux études antérieures traitant des trajectoires des gains des immigrants, et plus particulièrement en comparaison avec les dossiers des gains de la sécurité sociale utilisés aux États-Unis (Lubotsky, 2007). Tout d'abord, les données utilisées dans cette étude ne découlent pas d'un appariement de dossiers administratifs et de données d'enquête. Cela signifie que le biais potentiel découlant d'appariements non aléatoires ne représente pas une considération. Par ailleurs, les trajectoires des gains peuvent être comparées au moyen d'échantillons longitudinaux et quasi longitudinaux tirés de la même source de données; cela évite le problème de comparabilité qui se pose lorsque des ensembles de données différents sont utilisés. Un deuxième avantage est que les données sur les gains utilisées ici ne sont pas tronquées par le haut¹¹; cela évite les préoccupations liées à la troncation des valeurs supérieures de l'échantillon et aux changements connexes dans le plafond des gains au fil du temps. Enfin, les données utilisées dans la présente étude permettent de faire une distinction entre les immigrants établis et les travailleurs étrangers temporaires.

Néanmoins, la DAL-BDIM comporte des lacunes. De façon plus particulière, les données sur les gains longitudinaux sont disponibles uniquement pour les personnes qui produisent une déclaration de revenus (même si cela représente environ 95 % de la population en âge de travailler). Si aucune déclaration de revenus n'est observée, il n'est pas possible de déterminer si c'est parce que la personne n'a pas de travail, a quitté le Canada ou n'a pas produit de déclaration. Même si la DAL-BDIM comprend certains renseignements sur les caractéristiques des immigrants au moment de leur établissement au Canada, par exemple, leur niveau de scolarité et leur profession souhaitée, elle ne comprend pas de tels renseignements pour les personnes nées

8. La règle d'exclusion réelle est la suivante : *gains* > 500 \$CAN. Des vérifications de la robustesse ont été effectuées selon divers seuils (0 \$, 1 \$, 100 \$, 1 000 \$), sans effet sur les résultats.

9. La présente étude utilise les données du formulaire détaillé du recensement, qui est administré de façon aléatoire à 20 % de la population.

10. Par ailleurs, pour plus de cohérence, le présent document tient compte uniquement des immigrants qui sont arrivés au Canada à l'âge adulte (entre 25 et 44 ans).

11. La troncation par le haut se produit lorsque tous les gains au-dessus d'un niveau déterminé, par exemple 200 000 \$, sont déclarés comme étant à ce niveau.

au Canada. Les estimations des différences de gains peuvent donc être compromises par ce manque d'information. Cette lacune n'a pas de répercussions sur l'analyse présentée dans le présent document, toutefois, pour les raisons énoncées plus loin dans le document. Enfin, la BDIM comprend uniquement les immigrants qui se sont établis depuis 1980. Les personnes nées à l'étranger qui sont arrivées au Canada avant 1980 sont incluses dans le fichier, mais ne peuvent être « étiquetées » comme immigrantes. Ainsi, le groupe de comparaison ne comprend pas uniquement des personnes nées au Canada, mais aussi des immigrants qui sont arrivés avant 1980. Même si cela n'affecte pas les estimations des gains absolus des immigrants, les estimations de l'écart entre les gains des immigrants et ceux des personnes nées au Canada incluront les immigrants de « longue date » dans le groupe de comparaison.

L'ampleur de ce problème et la comparabilité de la DAL-BDIM et du recensement sont évaluées au tableau 6 en annexe. Pour que la comparaison soit possible, on ajoute, au groupe du recensement composé des personnes nées au Canada, les immigrants qui se sont établis avant 1980¹². Les résultats pour l'échantillon de la DAL-BDIM et l'échantillon du recensement montrent que les deux sources de données sont assez uniformes du point de vue de la taille absolue de la cohorte et de la proportion des cohortes constituées d'immigrants. Dans le groupe de comparaison du recensement, les données du recensement sont utilisées pour déterminer la proportion du groupe de comparaison constituée des *immigrants* de plus longue date, par opposition aux personnes nées au Canada. Pour la cohorte de 1990 à 1994, un peu plus de 3 % du groupe de comparaison est constitué d'immigrants de plus longue date (dans ce cas, ceux qui sont au Canada depuis 16 ans ou plus), et plus de 96 %, de personnes nées au Canada¹³. La proportion est plus grande dans la cohorte plus ancienne : un peu moins de 6 % du groupe de comparaison pour la cohorte de 1985 à 1989 est constitué d'immigrants qui sont arrivés au Canada avant 1980. Dans le cas de la cohorte la plus récente, la proportion d'immigrants dans le groupe de comparaison est négligeable.

Compte tenu des proportions relativement faibles du groupe de comparaison, qui sont constituées d'immigrants de plus longue date, et de la ressemblance économique de ces immigrants avec la population née au Canada, ce problème de groupe de comparaison n'est pas considéré comme particulièrement important¹⁴. Par ailleurs, le fait que l'étendue de la « contamination » varie d'une cohorte à l'autre ne pose pas de problème, étant donné que seules les comparaisons des trajectoires de l'écart des gains à l'intérieur des cohortes fondées sur des données longitudinales et transversales présentent un intérêt. Enfin, ce qui importe probablement le plus dans l'analyse empirique est la comparaison des résultats transversaux fondés sur les échantillons de la DAL-BDIM et les estimations obtenues du recensement. Dans la section qui suit, on montrera que les résultats du recensement, qui ne sont pas touchés par ce problème de groupe de comparaison, correspondent aux estimations découlant des données administratives.

4 Résultats empiriques

Les résultats empiriques commencent par certains modèles descriptifs fondés sur des données non corrigées (brutes). Dans les tableaux 1 et 2, on compare le niveau des gains des immigrants et l'écart entre les gains des immigrants et ceux des personnes nées au Canada, selon le nombre d'années écoulées depuis la migration, pour trois cohortes différentes, grâce à trois échantillons différents. Les trois échantillons sont les suivants : i) l'échantillon longitudinal de la DAL-BDIM;

12. On procède ainsi uniquement pour obtenir des statistiques descriptives dans le tableau 6 en annexe. Dans l'analyse empirique qui suit, les échantillons du recensement ne comprennent pas les immigrants établis avant 1980.

13. À noter que le groupe né au Canada comprend aussi les migrants enfants (de moins de 18 ans) qui sont arrivés au Canada avant 1980.

14. Par ailleurs, la majeure partie de l'attrition chez les immigrants se produit au cours des premières années au Canada (Aydemir et Robinson, 2008). Par conséquent, il est peu probable qu'une faible « contamination » par les immigrants de plus longue date ait des répercussions significatives sur les probabilités d'attrition dans le groupe de comparaison.

ii) l'échantillon transversal de la DAL-BDIM; et iii) l'échantillon transversal du recensement. Essentiellement, les tableaux 1 et 2 complètent les données comprises dans la figure 1.

Les différences entre les échantillons longitudinal et transversal de la DAL-BDIM (deux premières parties du tableau 1) sont au centre de la question. Pour toutes les cohortes, les gains des immigrants au cours de leurs toutes premières années au Canada ont tendance à être légèrement plus élevés dans l'échantillon longitudinal que dans l'échantillon transversal. À la fin de la période à l'étude (p. ex., après 20 ans passés au Canada pour la cohorte de 1985 et 1989), les gains sont identiques dans les deux échantillons. On veut obtenir ce résultat, étant donné que les deux échantillons sont identiques à ce moment-là et sont constitués de tous les immigrants qui étaient toujours au Canada et occupés après 20 ans. Toutefois, étant donné que les gains moyens sont légèrement plus faibles au début de la période et identiques à la fin, l'augmentation des gains est légèrement plus marquée dans l'échantillon transversal que dans l'échantillon longitudinal. Compte tenu de la discussion qui précède, c'est le résultat auquel on s'attendait. Du point de vue des écarts entre les gains des immigrants et ceux des personnes nées au Canada, toutefois, il y a peu de variation entre l'échantillon longitudinal et l'échantillon transversal. Ces tendances vont dans le sens de la constatation principale de l'analyse économétrique qui suit. Même s'il semble y avoir certaines différences dans l'augmentation des gains absolus des immigrants entre les échantillons transversal et longitudinal, par suite de l'attrition sélective dans les données transversales sur les immigrants, le même modèle d'attrition sélective est observé chez les travailleurs nés au Canada. Par conséquent, le biais découlant de cette attrition est annulé lorsque l'on compare les tendances des gains des immigrants et celles des personnes nées au Canada. Ainsi, l'écart entre les gains des immigrants et ceux des personnes nées au Canada diminue au fil du temps à un rythme similaire dans les deux échantillons.

La partie inférieure du tableau 1 comporte les mêmes statistiques pour les échantillons tirés du recensement. Lorsque l'on compare les cohortes pseudo-longitudinales élaborées à partir du recensement et des sections transversales de la DAL-BDIM, les trajectoires des gains sont très similaires. Pour toutes les cohortes, l'augmentation des gains des immigrants est à peu près la même, et il n'y a que des différences mineures dans la diminution des écarts entre les gains pour toutes les cohortes (voir le tableau 2). Les gains moyens des immigrants ont tendance à être plus élevés dans le recensement que dans les sections transversales de la DAL-BDIM. Cela va dans le sens des résultats de Frenette, Green et Picot (2006), qui ont déterminé que les valeurs de revenu de la moitié inférieure de la distribution sont plus faibles dans les données fiscales que dans les données du recensement. Tout bien considéré, les échantillons transversaux des deux sources de données produisent des trajectoires similaires des gains des immigrants, tant en termes absolus que relatifs (par rapport aux personnes nées au Canada).

Tableau 1
Gains moyens des immigrants, données longitudinales et données transversales

	Années de la mesure des gains			
	1990	1995	2000	2005
Données longitudinales dans la DAL-BDIM				
Cohorte de 1985 à 1989				
Taille de l'échantillon (nombre)	12 080	11 570	12 173	14 075
Gains (dollars)	32 100	38 200	48 600	49 100
Écart entre les gains (pourcentage)	-30,2	-22,8	-15,8	-14,8
Cohorte de 1990 à 1994				
Taille de l'échantillon (nombre)	...	19 832	21 496	24 934
Gains (dollars)	...	27 200	40 700	43 000
Écart entre les gains (pourcentage)	...	-40,4	-26,7	-25,6
Cohorte de 1995 à 1999				
Taille de l'échantillon (nombre)	21 278	24 902
Gains (dollars)	35 900	43 500
Écart entre les gains (pourcentage)	-30,0	-22,2
Section transversale répétée dans la DAL-BDIM				
Cohorte de 1985 à 1989				
Taille de l'échantillon (nombre)	19 484	15 567	14 637	14 075
Gains (dollars)	31 600	36 600	46 300	49 100
Écart entre les gains (pourcentage)	-29,0	-23,6	-16,6	-14,8
Cohorte de 1990 à 1994				
Taille de l'échantillon (nombre)	...	29 032	26 697	24 934
Gains (dollars)	...	26 100	38 950	43 000
Écart entre les gains (pourcentage)	...	-40,5	-27,6	-25,7
Cohorte de 1995 à 1999				
Taille de l'échantillon (nombre)	28 089	24 902
Gains (dollars)	34 100	43 500
Écart entre les gains (pourcentage)	-30,7	-22,1
Recensement				
Cohorte de 1985 à 1989				
Taille de l'échantillon (nombre)	20 746	17 625	16 645	16 493
Gains (dollars)	32 700	37 300	45 000	50 200
Écart entre les gains (pourcentage)	-24,2	-18,1	-12,1	-9,3
Cohorte de 1990 à 1994				
Taille de l'échantillon (nombre)	...	28 483	26 326	25 354
Gains (dollars)	...	27 600	39 100	45 300
Écart entre les gains (pourcentage)	...	-34,4	-21,4	-18,4
Cohorte de 1995 à 1999				
Taille de l'échantillon (nombre)	27 819	25 643
Gains (dollars)	36 000	45 200
Écart entre les gains (pourcentage)	-22,4	-16,2

Notes : Calculs des auteurs à partir de la DAL-BDIM et du Recensement de la population. La taille de l'échantillon a trait aux immigrants seulement. Chaque année, la population est constituée des hommes de 25 à 64 ans qui ont des gains positifs. Les immigrants avaient entre 25 et 44 ans lorsqu'ils ont migré. DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

Tableau 2**Augmentation des gains des immigrants et diminution de l'écart entre les gains, données longitudinales et données transversales**

Bases de données et cohortes	Augmentation des gains			Diminution de l'écart entre les gains		
	De 1990 à 2005	De 1995 à 2005	De 2000 à 2005	De 1990 à 2005	De 1995 à 2005	De 2000 à 2005
	pourcentage			points de pourcentage		
Données longitudinales dans la DAL-BDIM						
Cohorte de 1985 à 1989	53	15,4
Cohorte de 1990 à 1994	...	58	14,8	...
Cohorte de 1995 à 1998	21	7,8
Section transversale répétée dans la DAL-BDIM						
Cohorte de 1985 à 1989	55	14,2
Cohorte de 1990 à 1994	...	64	14,8	...
Cohorte de 1995 à 1998	27	8,6
Recensement						
Cohorte de 1985 à 1989	53	14,9
Cohorte de 1990 à 1994	...	64	15,9	...
Cohorte de 1995 à 1998	26	6,2

Notes : Calculs des auteurs à partir de la DAL-BDIM et du Recensement de la population. Chaque année, la population est constituée des hommes de 25 à 64 ans qui ont des gains positifs. Les immigrants avaient entre 25 et 44 ans lorsqu'ils ont migré. L'écart entre les gains représente la différence entre les gains des immigrants et les gains des personnes nées au Canada, divisée par les gains des personnes nées au Canada. DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

4.1 Estimations par régression

Les tendances qui précèdent sont fondées sur des données non corrigées (brutes), mais la plupart des résultats compris dans les ouvrages publiés antérieurement découlent d'une certaine forme de modèle de régression. On utilise, dans cette étude, un cadre économétrique standard pour examiner les gains absolus et relatifs des immigrants dans les données longitudinales et transversales répétées. Afin de ne pas imposer d'hypothèse d'augmentation constante des gains d'une cohorte à l'autre, les données des différentes cohortes ne sont pas regroupées; les trajectoires des gains des cohortes successives d'immigrants sont donc analysées séparément.

Les données recueillies aux États-Unis laissent supposer que la progression de la rémunération des immigrants dans les données longitudinales n'est pas uniforme d'une cohorte d'entrée à l'autre (Duleep et Regets, 1997; Duleep et Dowhan, 2002). Une part importante du biais estimé de « migration de sortie » dans Lubotsky (2007) semble découler uniquement de la cohorte de 1970 à 1979, et non pas des deux autres cohortes examinées. Les résultats précédents à partir des données non corrigées laissent supposer qu'il peut y avoir certaines différences entre les cohortes dans les données canadiennes aussi.

Même si la plupart des ouvrages empiriques portent sur l'augmentation des gains relatifs des immigrants, il est utile de décrire les trajectoires des niveaux des gains absolus des immigrants, ainsi que l'écart entre les gains des immigrants et ceux des personnes nées au Canada, et d'évaluer comment l'augmentation des gains et l'écart entre les gains diffèrent dans les données transversales et longitudinales. Ainsi, l'analyse commence par une estimation des trajectoires des gains absolus des nouveaux immigrants, des régressions distinctes étant exécutées pour chacune

des trois cohortes. Il existe une façon simple de rendre compte de ces tendances, à savoir estimer la régression suivante pour les cohortes d'entrée :

$$w_{it} = \alpha + \beta_1 \hat{Age}_{it} + \theta ysm_{it} + \varepsilon_{it} \quad (1)$$

où : w_{it} correspond au logarithme des gains annuels de la personne i à l'année t ; \hat{Age}_{it} est un polynôme en fonction de l'âge de la personne; et ysm_{it} , le nombre d'années passées dans le pays hôte depuis l'établissement, qui est précisé sous forme de variable catégorique (0 à 5 ans; 6 à 10 ans; 11 à 15 ans; et 16 à 20 ans). Dans cette régression touchant uniquement les immigrants, la colinéarité ne permet pas d'estimer d'effets de période; ainsi, aucun contrôle d'année civile n'est inclus. Le tableau 3 rend compte des coefficients estimés, au moyen du modèle 1, du nombre d'années écoulées depuis la migration pour les trois cohortes d'immigrants. Des estimations distinctes sont fournies pour les échantillons transversaux et longitudinaux de la DAL-BDIM, ainsi que pour l'échantillon du recensement.

Pour les trois cohortes, il existe des preuves que la trajectoire des gains des immigrants est surestimée dans l'échantillon transversal par rapport à l'échantillon longitudinal. Pour la cohorte de 1985 à 1989, le tableau 3 montre que l'augmentation des gains entre 0 et 5 ans et 16 et 20 ans au Canada se situait à 0,27 (environ 31 %) dans la DAL-BDIM longitudinale et à 0,33 (39 %) dans la DAL-BDIM transversale. C'est donc dire que l'augmentation des gains des immigrants après 16 à 20 ans au Canada diffère de 8 points de pourcentage. Ces résultats laissent supposer un biais à la hausse dans les résultats transversaux. Pour la cohorte de 1990 à 1994, l'augmentation des gains après 11 à 15 ans au Canada est de 0,39 dans la DAL-BDIM longitudinale et de 0,49 dans la DAL-BDIM transversale, une différence de 15 points de pourcentage. Un biais est aussi observé en ce qui a trait à la dernière cohorte (0,21 et 0,27). Il convient aussi de souligner que les estimations à partir de l'échantillon transversal de la DAL-BDIM correspondent dans une large mesure à celles fondées sur le recensement. Cela confirme que les données administratives fournissent des estimations des trajectoires des gains des immigrants similaires aux résultats publiés à partir des données du recensement.

De nombreux modèles standards de régression des gains tiennent compte de la scolarité; ainsi, le niveau de scolarité au moment de l'établissement des immigrants est ajouté à l'équation (1). Il en résulte qu'une partie du biais entre les échantillons longitudinal et transversal sera éliminé, si ce biais est le résultat d'une probabilité plus élevée de sortie parmi les immigrants moins scolarisés. Toutefois, cela n'est pas le cas. La différence d'augmentation des gains des immigrants estimée à partir des deux échantillons de la DAL-BDIM est à peu près la même, peu importe si la scolarité est prise en compte (tableaux 3 et 4)¹⁵. Toutefois, lorsque l'on examine l'augmentation des gains au fil du temps à l'intérieur des cohortes, on trouve des preuves d'augmentations plus marquées dans l'échantillon transversal que dans l'échantillon longitudinal, particulièrement pour la cohorte de 1990 à 1994. Cela fournit une preuve indirecte d'une probabilité plus élevée de sortie des immigrants à faibles gains à l'intérieur des groupes de niveaux de scolarité.

15. La présente étude ne comporte pas de régression conditionnelle fondée sur l'échantillon du recensement, les catégories de niveaux de scolarité utilisées dans la DAL-BDIM ne correspondant pas à celles comprises dans le recensement.

Tableau 3
Augmentation des gains des immigrants au Canada, données
longitudinales et données transversales

Nombre d'années écoulées depuis la migration	DAL-BDIM		Recensement
	Données longitudinales	Données transversales	
	Colonne 1	Colonne 2	Colonne 3
coefficient du modèle de régression			
Cohorte de 1985 à 1989			
6 à 10 ans	0,117	0,065	0,057
11 à 15 ans	0,271	0,266	0,237
16 à 20 ans	0,271	0,333	0,314
Cohorte de 1990 à 1994			
6 à 10 ans	0,321	0,371	0,374
11 à 15 ans	0,391	0,494	0,498
Cohorte de 1995 à 1999			
6 à 10 ans	0,216	0,275	0,267

Notes : Les coefficients fournissent une estimation approximative de l'augmentation en pourcentage; par exemple, l'augmentation des gains dans les données longitudinales entre 0 à 5 ans et 6 à 10 ans au Canada pour la cohorte arrivée entre 1985 et 1989 se situait à environ 11,7 %. La catégorie de référence est constituée des immigrants qui ont migré depuis 1 à 5 ans. Tous les coefficients sont statistiquement significatifs au niveau de 1 %. Le tableau rend compte des coefficients des variables fictives des « années écoulées depuis la migration » dans le modèle 1. DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

À partir des estimations des tableaux 3 et 4, on peut supposer que, parmi les immigrants des trois cohortes analysées, ceux qui sont sortis de l'échantillon étaient plus susceptibles d'avoir des gains plus faibles que ceux qui sont restés. Par conséquent, les changements au fil du temps dans la composition des sections transversales répétées découlant des sorties sélectives parmi les personnes ayant des gains plus faibles et plus élevés ont causé un biais dans les trajectoires des gains absolus estimées à partir des données transversales.

Toutefois, la plupart des recherches ne sont pas axées uniquement sur les trajectoires des gains des immigrants; elles portent plutôt sur les changements au fil du temps de l'écart entre les gains des immigrants et ceux des personnes nées au Canada. Pour tester un tel biais, les groupes de comparaison constitués des personnes nées au Canada décrits dans la section 3 sont ajoutés à l'analyse. Pour évaluer les augmentations des gains des immigrants (selon le nombre d'années écoulées depuis l'établissement) par rapport au groupe de comparaison, on applique un cadre empirique standard pour ce type d'analyse (Chiswick, 1978; Borjas, 1999).

Tableau 4

Augmentation des gains des immigrants au Canada (en tenant compte de la scolarité), données longitudinales et données transversales

Nombre d'années depuis la migration	DAL-BDIM	
	Données longitudinales	Données transversales
	coefficient du modèle de régression	
Cohorte de 1985 à 1989		
6 à 10 ans	0,143	0,096
11 à 15 ans	0,320	0,325
16 à 20 ans	0,335	0,408
Cohorte de 1990 à 1994		
6 à 10 ans	0,345	0,399
11 à 15 ans	0,435	0,536
Cohorte de 1995 à 1999		
6 à 10 ans	0,243	0,287

Notes : Les coefficients fournissent une estimation approximative de l'augmentation en pourcentage; par exemple, l'augmentation des gains dans les données longitudinales entre 0 à 5 ans et 6 à 10 ans au Canada pour la cohorte arrivée entre 1985 et 1989 se situait à environ 14,3 %. La catégorie de référence est constituée des immigrants qui ont migré depuis 1 à 5 ans. Tous les coefficients sont statistiquement significatifs au niveau de 5 %. Le tableau rend compte des coefficients des variables fictives des « années écoulées depuis la migration ». DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales et Base de données longitudinales sur les immigrants.

Prenons le modèle de régression suivant du logarithme des gains annuels.

$$w_{it} = \alpha + \beta_1 \hat{Age}_{it} + \beta_2 Année_{it} + \lambda I_i + \gamma M_i \cdot I_i + \theta ysm_{it} \cdot I_i + \varepsilon_{it} \quad (2)$$

où les variables additionnelles, outre celles de l'équation (1), sont les suivantes : un vecteur des variables fictives d'année civile, $Année_{it}$; l'âge de l'immigrant au moment de l'arrivée dans le pays hôte, M_i , afin d'obtenir une approximation de l'expérience sur le marché du travail étranger; et une variable fictive permettant de distinguer le statut d'immigrant d'avec celui de personne née au pays, I_i . À noter que dans le modèle (2), les variables M_i (âge des immigrants à l'arrivée dans le pays hôte) et ysm_{it} (nombre d'années écoulées depuis la migration) contiennent maintenant un terme d'interaction avec la variable de statut d'immigrant, permettant ainsi à la trajectoire des gains de différer entre les immigrants et les personnes nées au pays. Comme auparavant, un modèle distinct est exécuté pour chacune des trois cohortes.

Le coefficient des années écoulées depuis la migration, θ , est le paramètre d'intérêt et mesure la variation de l'écart entre les gains des immigrants et ceux des personnes nées au Canada selon le nombre d'années passées au Canada, ou, autrement dit, le taux de convergence des gains au fil du temps entre les immigrants et les personnes nées au Canada. L'écart entre les gains des immigrants au moment de l'arrivée et l'effet de l'expérience étrangère sur les gains dans le pays hôte sont rendus par λ et γ , respectivement. Afin de déterminer de façon distincte les coefficients des variables \hat{Age} , $Année$, M et ysm , on doit imposer la restriction selon laquelle les effets de l'âge et de la période, β_1 et β_2 , sont les mêmes pour les immigrants et les personnes nées au Canada. Comme l'expliquent Borjas (1999) et Lubotsky (2007), cette hypothèse n'est pas sans problème.

Toutefois, la plupart des estimations de l'augmentation des gains des immigrants sont fondées sur cette hypothèse, et elle est conservée ici pour évaluer la différence entre l'augmentation des gains mesurée à partir des échantillons longitudinal et transversal. Le modèle est le même que celui utilisé par Lubotsky (2007) aux États-Unis, auquel ces résultats peuvent être comparés¹⁶.

Les colonnes 1 et 2 du tableau 5 rendent compte des résultats de l'estimation de l'équation (2) basée sur l'échantillon longitudinal de la DAL-BDIM (colonne 1) et l'échantillon transversal de la DAL-BDIM (colonne 2). La colonne 3 montre les résultats à partir de l'échantillon du recensement. L'analyse commence par une comparaison des résultats sur la base des deux échantillons de la DAL-BDIM, éliminant ainsi toute différence attribuable aux sources de données (données d'enquêtes et données administratives). Pour la cohorte de 1985 à 1989, on ne note pas de différences majeures, que ce soit sur le plan de l'écart entre les gains initiaux ou sur le plan de la variation de l'écart entre les gains au fil du temps. Dans les deux échantillons, l'écart initial se situe autour de 0,33 à 0,35; après 16 à 20 ans au Canada, il diminue pour s'établir à environ 0,24 (environ 27 %). Pour la cohorte de 1990 à 1994, l'écart initial est légèrement plus élevé dans l'échantillon transversal (comme on devrait s'y attendre en présence d'un biais). Toutefois, la différence est statistiquement non significative, et la variation de l'écart ne diffère pas significativement d'un échantillon à l'autre, ce qui constitue l'intérêt principal de la présente étude. La même chose s'applique à la cohorte de 1995 à 1999. Globalement, les résultats découlant des données administratives ne laissent pas supposer l'existence d'un biais dans les estimations transversales répétées de l'assimilation des gains, contrairement à ce qui a été observé pour les gains absolus. On arrive à une conclusion similaire lorsque l'on compare les résultats longitudinaux de la DAL-BDIM et les résultats transversaux du recensement.

Un certain nombre de vérifications de la robustesse ont été effectuées. L'âge au moment de la migration a été élargi pour s'étendre de 20 à 54 ans, mais cela n'a pas modifié les résultats (voir les tableaux 7 et 8 en annexe). De même, la sensibilité des modèles à divers seuils de gains positifs (0 \$; 1 \$; 100 \$; 1 000 \$), ainsi qu'au retranchement des personnes dans les 1 % supérieurs/inférieurs de la distribution, a aussi été vérifiée; ni l'une ni l'autre de ces régressions additionnelles n'a eu de répercussions significatives¹⁷. Enfin, les modèles de régression (1) et (2) ont été estimés sur un échantillon regroupé de toutes les cohortes d'immigrants et de tous les groupes de comparaison. Ces modèles comprenaient des variables fictives de période d'arrivée et reposent sur le principe que la trajectoire des gains est la même pour toutes les cohortes. De telles estimations, même si elles nécessitent une hypothèse additionnelle, sont assez courantes dans les ouvrages publiés. Les coefficients estimés découlant de ces régressions, qui figurent dans les tableaux 9 et 10 en annexe, sont conformes aux résultats des régressions à l'intérieur des cohortes. C'est donc dire que les estimations longitudinales et transversales diffèrent dans le cas de l'augmentation des gains absolus des immigrants, mais pas pour l'écart entre les gains des immigrants et ceux des personnes nées au Canada.

16. Comme dans Lubotsky (2007), on ne tient pas compte du niveau de scolarité des personnes, l'objectif étant de vérifier s'il y a un biais dans les trajectoires des gains non-conditionnels des immigrants et des personnes nées au pays. La seule différence par rapport aux spécifications de Lubotsky concerne la variable de l'âge, pour laquelle il utilise plutôt une variable de l'expérience potentielle (âge moins nombre d'années écoulées depuis la fin des études). Les données sur le niveau de scolarité pour les personnes nées au Canada ne sont pas disponibles; par conséquent, aux fins du présent document, il n'est pas possible de faire une distinction entre les expériences potentielles sur le marché du travail.

17. Pour plus de concision, les résultats de ces régressions additionnelles ne sont pas fournis. Ils sont disponibles sur demande.

Tableau 5
Augmentation des gains relatifs des immigrants au Canada, données
longitudinales et données transversales

Nombre d'années écoulées depuis la migration	Variation dans l'écart entre les gains au fil du temps		
	DAL-BDIM		Recensement
	Données longitudinales	Données transversales	
	Colonne 1	Colonne 2	Colonne 3
coefficient du modèle de régression			
Cohorte de 1985 à 1989			
6 à 10 ans	0,121	0,081	0,094
11 à 15 ans	0,216	0,192	0,181
16 à 20 ans	0,247	0,242	0,239
0 an (écart initial)	-0,348	-0,328	-0,334
Cohorte de 1990 à 1994			
6 à 10 ans	0,212	0,221	0,236
11 à 15 ans	0,252	0,265	0,279
0 an (écart initial)	-0,447	-0,465	-0,444
Cohorte de 1995 à 1999			
6 à 10 ans	0,151	0,153	0,141
0 an (écart initial)	-0,267	-0,274	-0,224

Notes : Les coefficients fournissent une estimation approximative de l'augmentation en pourcentage; par exemple, l'augmentation des gains dans les données longitudinales entre 0 à 5 ans et 6 à 10 ans au Canada pour la cohorte arrivée entre 1985 et 1989 se situait à environ 12,1 %. La catégorie de référence est constituée des immigrants qui ont migré depuis 1 à 5 ans. Tous les coefficients sont statistiquement significatifs au niveau de 5 %. DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

D'une part, pris ensemble, les résultats empiriques de la présente étude ne fournissent pas de preuve que les estimations transversales des gains des immigrants par rapport à ceux des personnes nées au Canada sont biaisées de façon significative par l'attrition sélective, à tout le moins pour les trois cohortes d'immigrants incluses dans l'analyse. D'autre part, il existe des preuves que les estimations de l'augmentation des gains absolus des immigrants fondées sur des données transversales répétées du recensement sont légèrement biaisées à la hausse. Cela est probablement dû à une probabilité plus forte de sortie de l'échantillon des immigrants ayant des gains plus faibles. Pour mieux comprendre pourquoi il existe un biais dans la trajectoire des gains absolus, mais pas dans la trajectoire de l'écart entre les gains, les caractéristiques des gains des « personnes qui sortent » de l'échantillon et des « personnes qui restent » dans l'échantillon sont examinées plus étroitement.

5 Qui sort de l'échantillon?

La présente section vise à déterminer si les personnes qui étaient comprises dans l'échantillon administratif transversal, mais pas dans l'échantillon longitudinal, ont tendance de façon disproportionnée à être des personnes à faibles gains plutôt qu'à être réparties uniformément dans la distribution des gains. Ces personnes sont appelées « sortants ». Autrement dit, la probabilité de sortie de l'échantillon transversal est-elle plus grande pour les personnes ayant de faibles gains que pour celles qui ont des gains plus élevés? Si elle est plus grande pour les personnes ayant de

faibles gains, un tel modèle de sortie entraînera un biais à la hausse des trajectoires des gains sur la base des données transversales, du type observé précédemment pour l'augmentation des gains absolus, mais pas des gains relatifs.

Pour répondre à cette question, il est nécessaire de définir la distribution des gains et de déterminer où chacun des « sortants » se situe dans cette distribution, de la façon appropriée. Le principal problème a trait à la façon d'obtenir la distribution appropriée pour l'ensemble de la cohorte des immigrants, étant donné que certaines personnes quittent leur emploi tôt dans la période à l'étude, certaines quittent plus tard, et d'autres demeurent jusqu'à la fin, c'est-à-dire qu'elles « restent ».

On pourrait choisir la distribution des gains qui existait au cours des toutes premières années dans le pays hôte. Toutefois, la situation de l'immigrant dans la distribution changera probablement au cours des années ultérieures, lorsqu'il aura passé plus de temps dans le pays hôte. Par conséquent, une telle approche pourrait être trompeuse. Une deuxième option consisterait à calculer les gains moyens sur l'ensemble de la période à l'étude pour chaque individu et de fonder la distribution sur des moyennes individuelles. Toutefois, les gains augmentent au fur et à mesure des années écoulées depuis la migration. Ainsi, les gains de « sortants » précoces seraient sous-estimés de façon significative par rapport à ceux des personnes qui restent ou des « sortants » tardifs, et leur situation dans la distribution des gains serait déterminée de façon erronée. Ce qui nous intéresse, ce sont les résultats économiques des « sortants » pendant leurs années de travail au Canada, par rapport à tous les autres immigrants de la même cohorte au cours des mêmes années au Canada. À cette fin, pour une année donnée, T , une variable fictive de *sortie* est fixée à 1 pour toutes les personnes qui sont sorties de l'échantillon l'année T , mais qui ont été observées dans l'échantillon, l'année $T-1$. Les gains cumulatifs de chaque personne l'année T correspondent à la somme des gains de cette personne jusqu'à l'année précédente ($T-1$). Cela permet de définir une distribution des gains différente pour chaque année (p. ex., pour toutes les années entre 1986 et 2005 pour la cohorte de 1985), sur la base des gains cumulatifs entre 1986 et toute année donnée T . Le niveau des gains de chaque « sortant » par rapport à ceux de tous les autres immigrants de sa cohorte, pour les mêmes années, peut alors être déterminé. Cette approche est appliquée à l'année de la première sortie. Certains « sortants » peuvent être sortis de l'échantillon, être entrés à nouveau dans l'échantillon, puis sortis de l'échantillon, mais seule la première sortie est prise en compte ici.

Un modèle de régression simple est exécuté pour chaque décile de gains, afin d'estimer la probabilité de sortie, $Pr(exit = 1)$. À cette fin, on utilise un modèle probit dont la variable dépendante est la variable fictive de sortie définie précédemment. Les variables indépendantes comprennent des variables fictives pour les dix déciles de gains et un polynôme de deuxième degré dans l'âge. Les variables de contrôle de l'âge sont conçues pour éliminer tout effet du profil âge-gains sur la probabilité que les « sortants » appartiennent aux déciles de gains inférieurs¹⁸. Les données de la DAL-BDIM sont utilisées pour cet exercice, pour chacune des trois cohortes d'immigrants (1985, 1990 et 1995), ainsi que pour chaque groupe de comparaison respectif¹⁹.

Les résultats sont présentés aux graphiques 1 à 3. Pour chaque décile de gains, on trace la probabilité de sortie de l'échantillon transversal de la DAL-BDIM. À noter que la probabilité de sortie est toujours plus élevée pour les immigrants que pour les personnes nées au Canada. Une partie (ou la majorité) de cette différence pourrait être attribuable à une probabilité plus forte de migration de sortie du Canada pour les immigrants, mais les données ne permettent pas de différencier les raisons de la sortie de l'échantillon. Lorsque l'on examine les immigrants

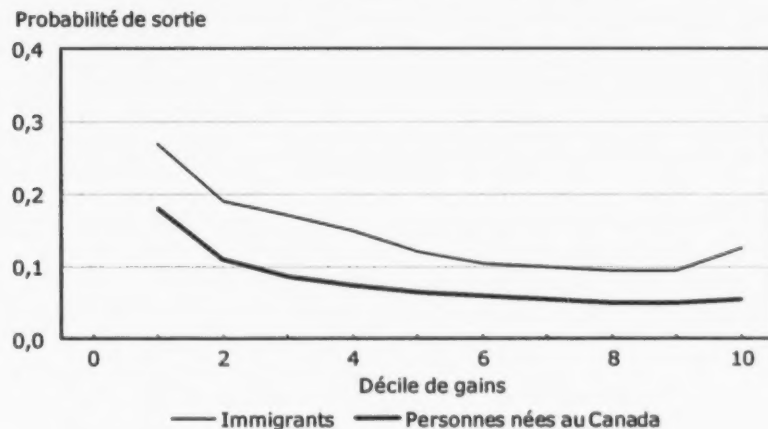
18. Les jeunes travailleurs ont tendance à avoir des gains plus faibles et à être plus mobiles que les travailleurs plus âgés.

19. Dans le présent document, on se penche sur des cohortes d'une année à la présente section, ce qui simplifie la détermination des déciles de gains cumulatifs. Les constatations se généralisent toutefois aux autres cohortes d'une année.

seulement, les résultats montrent une probabilité beaucoup plus grande de sortie de l'échantillon pour les immigrants des trois ou quatre déciles inférieurs que pour les immigrants des déciles supérieurs. Pour la cohorte de 1985, les immigrants du décile inférieur étaient susceptibles dans une proportion de près de 30 % de sortir de l'échantillon, comparativement à environ 10 % pour ceux se situant dans la partie supérieure de la distribution. Une tendance similaire est observée pour les deux autres cohortes. C'est cette tendance qui mène au biais à la hausse dans les gains absolus des immigrants observé dans les données transversales répétées²⁰.

Graphique 1

Probabilité de sortie par décile de gains — Cohorte d'immigrants de 1985 et groupe de comparaison



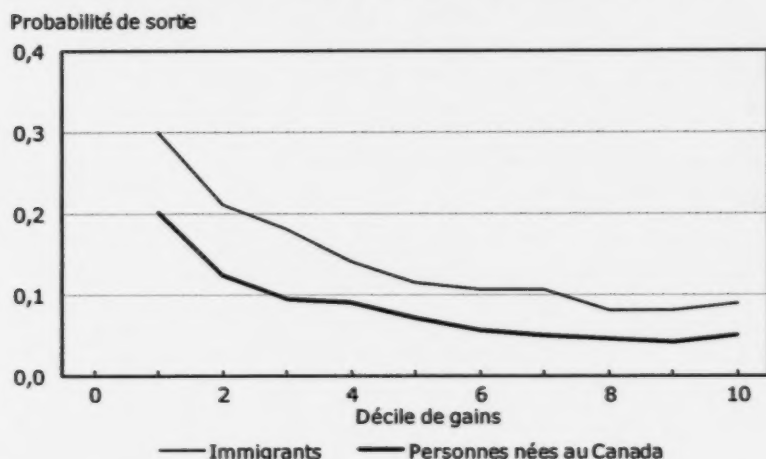
Note : Estimation des auteurs.

Sources : Statistique Canada, banque de Données administratives longitudinales et Base de données longitudinales sur les immigrants.

20. Une tendance intéressante à noter est l'augmentation de la probabilité de sortie au sommet de la distribution des gains pour les immigrants, particulièrement dans les cas de la cohorte de 1985. Cela pourrait être relié au phénomène des immigrants très qualifiés qui utilisent le Canada comme porte d'entrée sur le marché du travail de l'Amérique du Nord.

Graphique 2

Probabilité de sortie par décile de gains — Cohorte d'immigrants de 1990 et groupe de comparaison

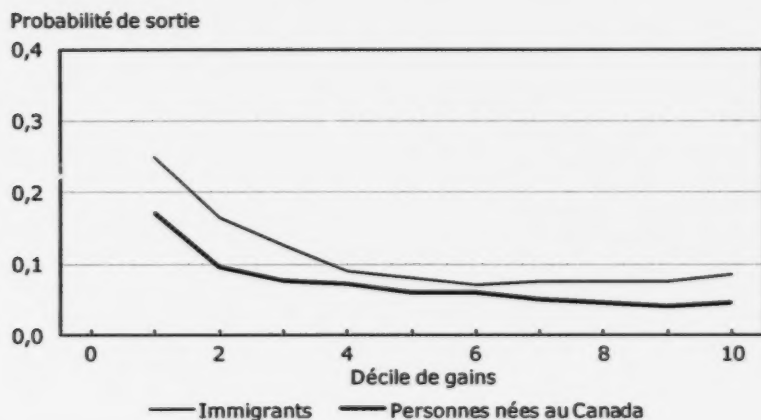


Note : Estimation des auteurs.

Sources : Statistique Canada, banque de Données administratives longitudinales et Base de données longitudinales sur les immigrants.

Graphique 3

Probabilité de sortie par décile de gains — Cohorte d'immigrants de 1995 et groupe de comparaison



Note : Estimation des auteurs.

Sources : Statistique Canada, banque de Données administratives longitudinales et Base de données longitudinales sur les immigrants.

Le degré du biais pour toute cohorte dépend de la différence de probabilité de sortie entre ceux qui ont de faibles gains et ceux qui ont des gains élevés; cette différence varie d'une cohorte à l'autre. Pour la cohorte de 1985, la probabilité de sortie dans les trois déciles inférieurs (à environ 20 %) représente 1,9 fois celle des trois déciles supérieurs (à environ 11 %). Dans la cohorte de 1995, le ratio est de 2,2. Les différences sont les plus grandes pour la cohorte de 1990, et la probabilité de sortie est 2,7 fois plus élevée parmi les personnes qui ont de faibles gains (trois déciles inférieurs) que parmi celles qui ont des gains élevés. Cela correspond au biais à la hausse plus grand dans

les trajectoires des gains absolus de la cohorte de 1995 comparativement aux autres cohortes, dont les trajectoires sont estimées à partir des données transversales.

Qu'en est-il du fait que peu de biais ait été observé dans les trajectoires des gains relatifs dans les données transversales? Lorsque l'on se penche sur le groupe de comparaison, on observe à peu près les mêmes tendances : la probabilité de sortie est significativement plus élevée parmi les travailleurs au bas de la distribution des gains que parmi ceux au sommet. Cela explique pourquoi le biais n'est pas observé dans l'écart entre les gains des immigrants et ceux des personnes nées au Canada au fil du temps. La probabilité de sortie est plus faible chez les personnes nées au Canada que chez les immigrants à tous les points de la distribution. Cela n'est pas étonnant, les immigrants étant plus susceptibles de cesser d'avoir des gains que les personnes nées au Canada, particulièrement au cours de leurs premières années au Canada, lorsque leur stabilité d'emploi est moins grande. Toutefois, la principale préoccupation à cet égard a trait à la probabilité relative de sortie des personnes qui ont de faibles gains ou des gains élevés. Dans ce cas, on peut observer la même tendance chez les personnes nées au Canada que chez les immigrants. L'augmentation des gains des cohortes d'immigrants et ceux des groupes de comparaison estimée à partir des données transversales est biaisée à la hausse, du fait de la probabilité plus grande que les personnes ayant des gains plus faibles quittent l'échantillon. L'écart, qui représente la différence entre ces deux trajectoires, n'est par conséquent que peu touché, le biais semblant s'appliquer à la fois aux immigrants et aux personnes nées au Canada de la même façon²¹.

6 Conclusions

La majorité des études traitant des trajectoires des gains des immigrants ont été fondées sur des données du recensement et ont utilisé des cohortes pseudo-longitudinales élaborées à partir de sections transversales répétées. Toutefois, la population de ces types de données change au fil du temps, certains membres de la cohorte des immigrants quittant le pays ou, à tout le moins, leur emploi. Si un effet de sélection découle de la sortie disproportionnée des immigrants qui obtiennent moins de succès et qui sont moins bien rémunérés (comparativement aux immigrants ayant des gains plus élevés), l'augmentation des gains estimée à partir des données transversales répétées sera biaisée à la hausse. Une partie du progrès des gains des immigrants ne sera pas « réelle », mais sera plutôt attribuable au changement de composition de la cohorte des immigrants au fil du temps.

Dans la présente étude, on utilise des données administratives longitudinales et des données du recensement pour déterminer si un tel biais existe au Canada. On se penche à la fois sur les trajectoires des gains absolus des cohortes de nouveaux immigrants et les trajectoires de l'écart entre les gains des immigrants et ceux des personnes nées au Canada au cours de leur 20 premières années passées au Canada. En ce qui a trait à l'augmentation des gains absolus des cohortes d'immigrants, les résultats fondés sur une section transversale répétée diffèrent des résultats découlant de données longitudinales. Dans les trois cohortes d'immigrants examinées, l'augmentation des gains (en fonction du nombre d'années passées au Canada), sur la base d'un modèle de régression standard, était de 7 points de pourcentage à 15 points de pourcentage plus élevé dans l'échantillon transversal répété que dans l'échantillon longitudinal. Ce biais a été causé parce que les immigrants de cette cohorte qui se situaient dans le tiers inférieur de la distribution des gains étaient au moins deux fois plus susceptibles de sortir de l'échantillon que ceux qui se situaient dans le tiers supérieur.

21. Les résultats de la présente étude ne varient pas lorsque l'on utilise un modèle de probabilité logit ou linéaire plutôt qu'un modèle probit. Les résultats sont aussi robustes à une autre définition des gains cumulatifs qui ne tient pas compte des gains de l'année précédant la sortie (afin d'éviter les problèmes possibles découlant de données sur les gains pour des parties d'année).

Toutefois, la variation des gains des immigrants *par rapport* aux gains des personnes nées au Canada, qui est le thème de la plupart des ouvrages publiés dans ce domaine, est très similaire dans les échantillons transversal et longitudinal. Il existe donc peu de preuve d'un biais important à la hausse dans la trajectoire de l'écart *entre les gains*. Même si les immigrants qui réussissent moins bien et qui sont moins bien rémunérés dans les diverses cohortes sont plus susceptibles de sortir de l'échantillon, il en va de même pour les personnes nées au Canada. C'est donc dire que l'augmentation des gains des échantillons d'immigrants et de personnes nées au Canada est surestimée dans les données transversales, à peu près dans la même mesure. C'est pourquoi, également, la trajectoire de l'écart entre les gains estimée au moyen de données longitudinales et d'un modèle standard des gains ne fait pas ressortir de biais dans les études antérieures des gains des immigrants au Canada.

Une comparaison de ces résultats et de ceux obtenus aux États-Unis (Lubotsky, 2007) montre que les immigrants au Canada affichent des modèles de participation au marché du travail qui s'apparentent davantage à ceux des personnes nées au pays que ce n'est le cas aux États-Unis.

7 Annexe

Tableau 6
Fréquence des immigrants au fil du temps dans les données du recensement et de la DAL-BDIM

Année	Échantillon de la DAL-BDIM			Échantillon du recensement		Groupe de comparaison du recensement	
	Taille de l'échantillon	Immigrants		Taille de l'échantillon	Immigrants	Immigrants qui se sont établis avant 1980	
	nombre	pourcentage		nombre	pourcentage		
Cohorte de 1985 à 1989							
1990	801 772	19 494	2,43	863 459	20 746	2,40	5,98
1995	748 578	15 576	2,08	803 699	17 625	2,19	5,72
2000	710 157	14 689	2,07	749 061	16 645	2,22	5,72
2005	663 986	14 082	2,12	697 100	16 493	2,37	5,56
Cohorte de 1990 à 1994							
1995	804 403	29 049	3,61	856 416	28 483	3,33	3,27
2000	776 789	26 703	3,44	815 563	26 326	3,23	3,31
2005	753 034	24 940	3,31	782 863	25 354	3,24	3,33
Cohorte de 1995 à 1999							
2000	780 890	28 116	3,60	813 354	27 819	3,42	1,28
2005	770 008	24 925	3,24	795 307	25 643	3,22	1,34

Notes : Calculs des auteurs à partir des données de la DAL-BDIM et du Recensement de la population. Chaque année, les observations se limitent aux personnes de 25 à 64 ans ayant des gains positifs. Les immigrants avaient entre 25 et 44 ans lorsqu'ils ont migré. Le *groupe de comparaison* a trait à l'échantillon utilisé dans l'étude en tant que groupe de personnes nées au Canada dans les fichiers de la DAL-BDIM (qui comprend les immigrants établis avant 1980). DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

Tableau 7

**Test de sensibilité pour l'âge au moment de la migration —
Augmentation des gains absolus des immigrants**

Nombre d'années écoulées depuis la migration	DAL-BDIM		Recensement
	Données longitudinales	Données transversales	
	Colonne 1	Colonne 2	Colonne 3
coefficient du modèle de régression			
Cohorte de 1965 à 1989			
6 à 10 ans	0,103	0,056	0,036
11 à 15 ans	0,251	0,255	0,236
16 à 20 ans	0,248	0,330	0,311
Cohorte de 1990 à 1994			
6 à 10 ans	0,323	0,360	0,359
11 à 15 ans	0,399	0,491	0,481
Cohorte de 1995 à 1999			
6 à 10 ans	0,206	0,258	0,233

Notes : La catégorie de référence est constituée des immigrants âgés de 20 à 54 ans au moment de la migration. Tous les coefficients sont statistiquement significatifs au niveau de 1 %. Le tableau rend compte des coefficients des variables fictives du « nombre d'années écoulées depuis la migration » dans la colonne 1. DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

Tableau 8
Test de sensibilité pour l'âge au moment de la migration —
Augmentation des gains relatifs des immigrants

Nombre d'années écoulées depuis la migration	Variation de l'écart entre les gains au fil du temps		
	DAL-BDIM		Recensement
	Données longitudinales	Données transversales	
	Colonne 1	Colonne 2	Colonne 3
	coefficient du modèle de régression		
Cohorte de 1985 à 1989			
6 à 10 ans	0,111	0,084	0,082
11 à 15 ans	0,195	0,187	0,181
16 à 20 ans	0,215	0,222	0,223
0 an (écart initial)	-0,333	-0,323	-0,327
Cohorte de 1990 à 1994			
6 à 10 ans	0,201	0,216	0,218
11 à 15 ans	0,247	0,265	0,276
0 an (écart initial)	-0,447	-0,469	-0,461
Cohorte de 1995 à 1999			
6 à 10 ans	0,148	0,151	0,144
0 an (écart initial)	-0,264	-0,266	-0,211

Notes : La catégorie de référence est constituée des immigrants âgés de 20 à 54 ans au moment de la migration. Tous les coefficients sont statistiquement significatifs au niveau de 5 %. Le tableau rend compte des coefficients des variables fictives du « nombre d'années écoulées depuis la migration ». DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

Tableau 9
Augmentation des gains absolus des immigrants d'une cohorte à
l'autre (échantillon regroupé)

Nombre d'années écoulées depuis la migration	DAL-BDIM		Recensement
	Données longitudinales	Données transversales	
	Colonne 1	Colonne 2	Colonne 3
	coefficient du modèle de régression		
6 à 10 ans	0,236	0,259	0,253
11 à 15 ans	0,360	0,423	0,410
16 à 20 ans	0,391	0,495	0,476

Notes : La catégorie de référence est constituée des immigrants qui ont migré depuis 1 à 5 ans. Tous les coefficients sont statistiquement significatifs au niveau de 1 %. DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

Tableau 10
Augmentation des gains relatifs des immigrants d'une cohorte à l'autre (échantillon regroupé)

	Variation de l'écart entre les gains au fil du temps		
	DAL-BDIM		Recensement
	Données longitudinales	Données transversales	
	Colonne 1	Colonne 2	Colonne 3
	coefficient du modèle de régression		
Nombre d'années écoulées depuis la migration			
6 à 10 ans	0,171	0,164	0,168
11 à 15 ans	0,234	0,233	0,232
16 à 20 ans	0,264	0,277	0,269
Variables fictives de cohorte			
Cohorte de 1985 à 1989	-0,323	-0,326	-0,332
Cohorte de 1990 à 1994	-0,438	-0,446	-0,415
Cohorte de 1995 à 1999	-0,329	-0,321	-0,285

Notes : Les catégories de référence sont constituées des immigrants qui ont migré depuis 1 à 5 ans. Tous les coefficients sont statistiquement significatifs au niveau de 1 %. DAL : banque de Données administratives longitudinales; BDIM : Base de données longitudinales sur les immigrants.

Sources : Statistique Canada, banque de Données administratives longitudinales, Base de données longitudinales sur les immigrants et Recensement de la population.

Bibliographie

Aydemir, A., et C. Robinson. 2008. « Global labour markets, return, and onward migration ». *Canadian Journal of Economics*. Vol. 41. N° 4. p. 1285 à 1311.

Baker, M., et D. Benjamin. 1994. « The performance of immigrants in the Canadian labor market ». *Journal of Labor Economics*. Vol. 12. N° 3. p. 369 à 405.

Bloom, D.E., et M. Gunderson. 1991. « An analysis of the earnings of Canadian immigrants ». *Immigration, Trade, and the Labor Market*. Publié sous la direction de J.M. Abowd et R.B. Freeman. Chicago. The University of Chicago Press.

Borjas, G.J. 1999. « The economic analysis of immigration ». *Handbook of Labor Economics*. Vol. 3. Publié sous la direction de D. Card et O. Ashenfelter. Amsterdam. North-Holland.

Chiswick, B.R. 1978. « The effect of Americanization on the earnings of foreign-born men ». *Journal of Political Economy*. Vol. 86. N° 5. p. 897 à 921.

Duleep, H.O., et D. Dowhan. 2002. « Insights from longitudinal data on the earnings growth of U.S. foreign-born men ». *Demography*. Vol. 39. N° 3. p. 485 à 506.

Duleep, H.O., et M. Regets. 1997. « Measuring immigrant wage growth using matched CPS files ». *Demography*. Vol. 34. N° 2. p. 239 à 249.

Edin, P.-A., R. LaLonde, et O. Åslund. 2000. « Emigration of immigrants and measures of immigrant assimilation: Evidence from Sweden ». *Swedish Economic Policy Review*. Vol. 7. p. 163 à 204.

Frenette, M., D.A. Green, et G. Picot. 2006. « Rising income inequality in the 1990s: An exploration of three data sources ». *Dimensions of Inequality in Canada*. Publié sous la direction de D.A. Green et J.R. Kesselman. Vancouver, Colombie-Britannique. University of British Columbia Press.

Frenette, M., et R. Morissette. 2003. *Convergeront-ils un jour? Les gains des travailleurs immigrants et de ceux nés au Canada au cours des dernières décennies*. Produit n° 11F0019M au catalogue de Statistique Canada. Ottawa, Ontario. Direction des études analytiques : documents de recherche. N° 215.

Grant, M.L. 1999. « Evidence of new immigrant assimilation in Canada ». *Canadian Journal of Economics*. Vol. 32. N° 4. p. 930 à 955.

Green, D.A., et C. Worswick. 2004. *Immigrant Earnings Profiles in the Presence of Human Capital Investment: Measuring Cohort and Macro Effects*. Polycopié.

Hu, W.-Y. 2000. « Immigrant earnings assimilation: Estimates from longitudinal data ». *American Economic Review Papers and Proceedings*. Vol. 90. N° 2. p. 368 à 372.

Hum, D., et W. Simpson. 2000. « Closing the wage gap: Economic assimilation of Canadian immigrants reconsidered ». *Journal of International Migrations and Integration / Revue de l'intégration et de la migration internationale*. Vol. 1. N° 4. p. 427 à 441.

Hum, D., et W. Simpson. 2004. « Reinterpreting the performance of immigrant wages from panel data ». *Empirical Economics*. Vol. 29. N° 1. p. 129 à 147.

Lubotsky, D. 2007. « Chutes or ladders? A longitudinal analysis of immigrant earnings ». *Journal of Political Economy*. Vol. 115. N° 5. p. 820 à 867.

McDonald, J.T., et C. Worswick. 1998. « The earnings of immigrant men in Canada: Job tenure, cohort, and macroeconomic conditions ». *Industrial & Labor Relations Review*. Vol. 51. N° 3. p. 465 à 482.

Meng, R. 1987. « The earnings of Canadian immigrant and native-born males ». *Applied Economics*. Vol. 19. N° 8. p. 1107 à 1119.

Picot, G., et A. Sweetman. 2005. *Dégradation du bien-être économique des immigrants et causes possibles : mise à jour 2005*. Produit n° 11F0019M au catalogue de Statistique Canada. Ottawa, Ontario. Direction des études analytiques : documents de recherche. N° 262.

Skuterud, M., et M. Su. 2009. *Immigrant Wage Assimilation and the Return to Foreign and Host-Country Sources of Human Capital* [Assimilation salariale des immigrants et rentabilité des sources étrangères et nationales de capital humain]. Réseau Canadien de Chercheurs dans le domaine du Marché du Travail et des Compétences (RCCMTC). Rapport de recherche. N° 30. Résumé en français.

Statistique Canada. 2009. *Dictionnaire des données administratives longitudinales, 2006*. Produit n° 12-585-X au catalogue de Statistique Canada. Ottawa, Ontario.